AD_____

GRANT NUMBER DAMD17-97-1-7202

TITLE: Investigation of Genetic Algorithms for Computer-Aided Diagnosis

PRINCIPAL INVESTIGATOR: Matthew A. Kupinski

CONTRACTING ORGANIZATION: The University of Chicago
Chicago, Illinois 60637

REPORT DATE: October 1998

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1998	Annual (1 Oct 97 - 30 Sep 98)

4. TITLE AND SUBTITLE

Investigation of Genetic Algorithms for Computer-Aided Diagnosis

5. FUNDING NUMBERS

DAMD17-97-1-7202

6. AUTHOR(S)

Kupinski, Matthew A.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The University of Chicago
Chicago, Illinois 60637

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Computer-aided diagnosis has the potential of substantially increasing diagnostic accuracy in mammography. Using a computer to double-check a radiologist's findings is becoming more popular and more important as the public learns that the best defense against breast cancer is early detection. The University of Chicago is currently developing computerized schemes to detect cancers in digital mammograms. We use a pattern recognition system known as an artificial neural network (ANN) to classify certain regions of the digital mammograms as cancerous or non-cancerous. ANNs are trained pattern recognition devices that take, as inputs, features extracted from regions in the mammograms and output the classification. Currently, there are a total of 71 features extracted from the various regions in each mammogram. A subset of those 71 features must be chosen as inputs for the ANN. The goal of the proposed research is to apply a technique known as a genetic algorithm and other optimization techniques to find the subset of features which would result in the best ANN performance. By improving the inputs to the ANN, the performance of the neural network and hence, the performance of the mass CAD scheme, should improve. Preliminary results have exhibited this improvement.

14. SUBJECT TERMS

Breast Cancer Computer-aided Diagnosis, Feature Selection, Artificial Neural Networks, Genetic Algorithms.

15. NUMBER OF PAGES 27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

_____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

_____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

_____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

_____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

_____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Matthew A Krupnick 10-19-98

PI - Signature Date

Contents

4 Introduction

4.1 Nature of the Problem

Breast cancer is a major cause of death among women over the age of forty [1]. Mammography is the most effective diagnostic procedure for the early detection of breast cancer [2, 3]. Mammography is not, however, perfect. Between 10-30% of women who have breast cancer and undergo mammography have negative mammograms [4–7]. Of these, radiologists have determined, retrospectively, that two-thirds of the cancers could have been detected [5, 6, 8, 9]. One possible means by which to decrease this number is to have two radiologists read the mammograms. This method has been shown to increase sensitivity by as much as 15%, [10, 11] but can be costly both financially and with respect to time. A computer-aided diagnostic scheme may act as an inexpensive second reading method. The final decision would be made by the radiologist. One current method being studied locates potential lesions by bilateral subtraction of images of the left and right breasts [12–14]. This method is based on the deviation from the normal architectural symmetry of the left and right breasts, with asymmetries corresponding to potential masses. The images are aligned and then non-linearly subtracted to create a run length image that enhances regions of potential lesions. These regions of interest (ROIs) are subsequently sent through feature analysis. Features from these potential lesions are extracted for input into an artificial neural network (ANN) where the decision of whether the ROI is a lesion or not is made.

The proposed research seeks to answer questions that arise when using artificial neural networks in decision making applications. Problems occur when the number of inputs used in the ANN become large. The development of a systematic method for determining the optimal subset of features to use must be developed. For this reason, genetic algorithms are currently being studied to alleviate this problem. Genetic algorithms may have the ability to optimize the inputs used in a ANN. Because neural networks play such a vital role in decreasing the number of false-positive detections, these genetic algorithms may dramatically improving the performance of the ANN and, hence, the overall performance of the CAD scheme. When successful, this technique will have wide ranging benefits to other mammography CAD schemes as well as many different applications of neural networks in decision making situations.

4.2 Background

Artificial neural networks (ANNs) are powerful pattern recognition systems. They differ from conventional algorithmic approaches to pattern recognition in that they do

not use pre-defined rules for categorizing data. Instead, ANNs learn from examples that are presented repeatedly. Neural networks have found increasing popularity in many different fields due to their ability to made decisions or draw conclusions based on complex, noisy or incomplete data. ANNs are also capable of processing large amounts of data quickly and are therefore usually more efficient than other methods.

Recently, neural networks have been applied to the field of computer-aided diagnostic imaging [15]. These applications include the diagnosis of masses in digital mammograms [16–20]. Artificial neural networks are part of a computer aided diagnostic (CAD) scheme being developed at the Kurt Rossmann Laboratory at the University of Chicago to detect lesions in digital mammograms thus providing a second opinion to radiologists.

Despite the classification power of artificial neural networks, problems in training can arise when the ANN structure becomes too complex or when the features selected for input do not combine to improve the separation function learned by the ANN [21]. Hence, when the number of possible inputs or features becomes large, a search technique should be applied to select those features which will result in the best ANN performance.

A genetic algorithm is a search technique loosely based on the principles of genetic variation and natural selection. Genetic algorithms are of particular interest because of their ability to find solutions to problems contained in enormous and complex search spaces [22]. They have provided solutions to a wide variety of problems in function optimization, [23, 24] image processing, [25, 26] and analysis of physical systems [24] to name a few.

Genetic algorithms are based on evolution. Potential solutions to problems are subjected to an artificial environment which promotes the survival of individual solutions which closely approximate the solution sought. These fittest potential solutions win the right to carry on to the next generation, exchange data with other potential solutions or be subject to mutation. This survival-of-the-fittest strategy usually results in the rapid approximation of the solution to the problem defined.

Receiver operating characteristic (ROC) analysis [27,28] will be employed to evaluate the performance of the ANN, and hence the performance of the genetic algorithm, in distinguishing true lesions from false-positives. The LABROC4 program developed by Metz *et al.* [29] will be used to fit the data output from the neural networks. The area, A_z, under the ROC curve represents the performance of the ANN. Free-response operating characteristic (FROC) curves, obtained by plotting the sensitivity (lesions detected divided by the actual number of lesions) versus the number of false positives per image, will also used.

4.3 Purpose

The purpose of this proposed research is to improve the performance of the mass CAD scheme by optimizing the subset of input features used by the artificial neural network. A genetic algorithm should provide the basis for this optimization and has the potential of greatly improving the overall performance of the CAD scheme.

5 Body

5.1 Technical Objectives

The objectives of this project are as follows:

- Development of a genetic algorithm for the optimization of artificial neural network inputs.
- Comparison of a genetic algorithm with other selection and optimization methods including previously used selection methods.
- Analysis of features selected by genetic algorithm and comparison of those features with visual techniques employed by radiologists.
- Development of a parallel genetic algorithm to improve performance of the search and to provide an even greater performance increase to the mass detection CAD program.

5.2 Methods

5.2.1 Development of Genetic Algorithm

The foundation of the genetic algorithm is the genetic representation of a solution to a defined problem. The most common and usually the most effective method employs representation of the solution as a binary string. This is also known as a string with a binary cardinality. For the project proposed, the cardinality is the total number of features to be sampled during the GA's run. Not only must the solution be represented as a string but that string, or solution, must have a performance value, or fitness, associated with it. In the case of this project that performance value would represent an ANN performance or A_z obtained from ROC analysis. An example of such a string and fitness is as follows:

```
Solution 1:  1 42 12 31 22 65 :  Fitness 0.97
```

In this example, the solution says that the first, forty-second, etc. features were used as input for the ANN and that the A_z, or performance of the ANN, was 0.97.

Figure 1 represents a schematic view of a genetic algorithm. First, a completely random set of strings are created for the initial generation. The fitnesses of these

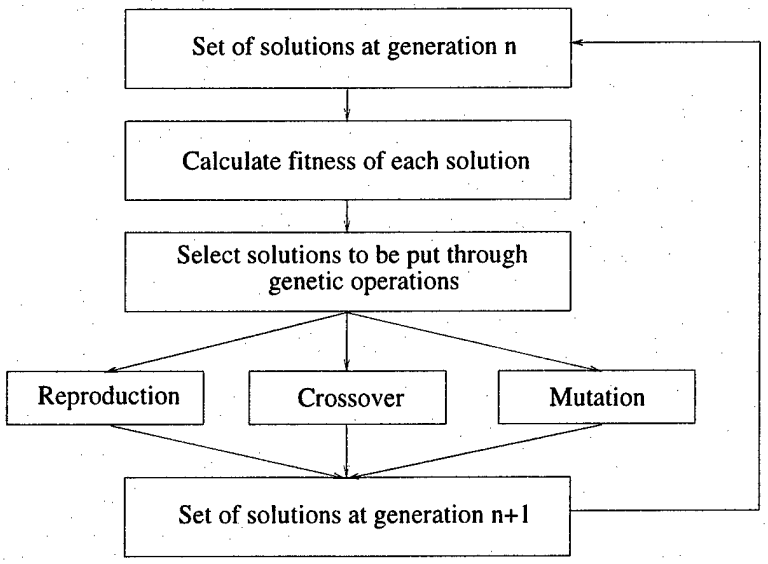

Figure 1: Schematic view of a genetic algorithm.

strings are calculated. Each string has a probability, based on its fitness value, of being selected for a genetic operator. Each operator also has a probability of occurrence. The three main genetic operators are reproduction, mutation and crossover. A genetic operator takes a string or strings, possibly modifies the string and places it in the next generation. Once the sting or strings have been selected and a genetic operator has been selected as well, then the operated string is placed in the next generation. This method continues until a completely new generation of solutions is present and the process starts over again from this new set of strings.

As alluded to earlier, the main focus of this research will be to apply a genetic algorithm to optimize the subset of features used as inputs to the ANN. The problem is that there are a total of 91 features and a subset of around 10-15 features must be selected. This means that there are on the order of 1016 combinations to choose from. This enormous search space and the fact that there is little known about the search space led to the conclusion that this was a problem suitable for a genetic algorithm.

The premise is that each string will represent a set of input features and the fitness of each string will be defined as the performance of the ANN with that set of input features. All the processes, i.e., mutation, selection and crossover, will be applied in a manner similar to that described above in an effort to find a set of input features that improves previous results, which were achieved using one-dimensional separation analysis. An effective method would be to use round robin outputs from a three-layered ANN like the one used in the mass detection scheme. This, however, is not practical. The typical round robin run can take more than 30 hours to complete 300

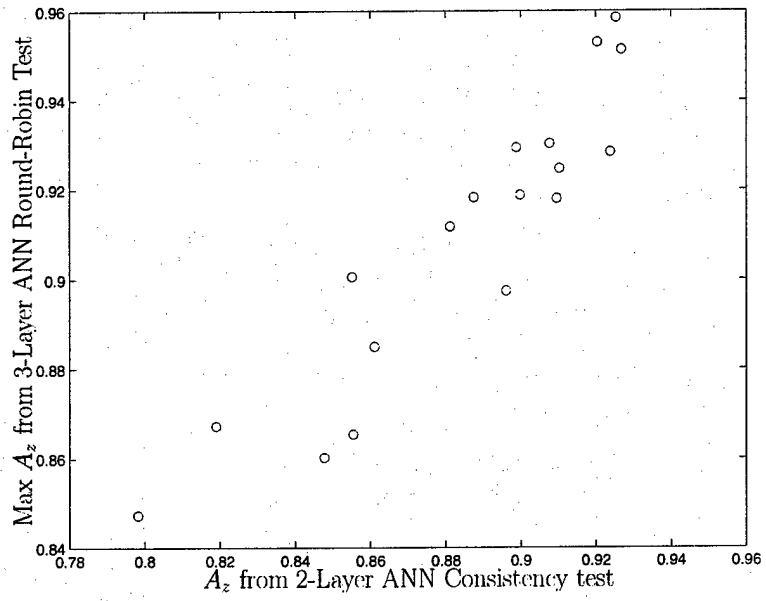

Figure 2: Plot of the 2-layer consistency A_z versus the 3-layer round robin A_z for different feature sets. The monotonic relationship show that the 2-layer consistency A_z can be used as the fitness function.

iterations. The typical GA with a population of 20 which runs for 1000 generations would not finish in 20 years if this round-robin, 3-layered network were employed. It is also not practical to use consistency outputs from the 3-layered network. These take much less time but a perfect result ($A_z = 1.0$) for consistency is common, so the better sets of features are indistinguishable from other sets. It was discovered, however, that there was a positive correlation between the consistency A_z of a two-layered (linear) ANN, which takes very little time to run, and the round robin A_z of a three- layered ANN using those same inputs (see Figure 2). This indicates that if the linear consistency A_z is high for a set of inputs, the non-linear round robin A_z will also be high. It does not, however, mean that, if there is a high non-linear round robin A_z, the linear consistency will be high as well. The major drawback to the use of this method is that it tends to limit the input sets to those that will combine and have excellent linear separation and ignores those sets which may combine to have an excellent performance with a highly non-linear separation.

The initial results of a 1000 generation run are shown in Figure 3 (labeled "Proportional Biasing"). As this figure indicates, there are areas where the performance does go up, but the overall average fitness is not impressive. The problem lies in the proportional biasing of the selection process based solely on the fitness. The are many sets of inputs that will have an A_z of about 0.95. There are, however, far fewer sets that will have an A_z of 0.96. The difficulty in going from a 0.95 to a 0.96 is not

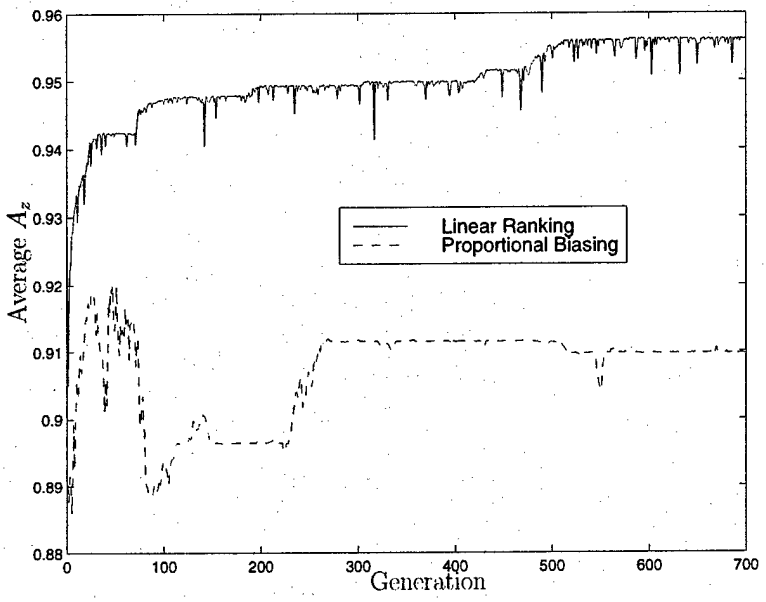

Figure 3: Performance of genetic algorithm using proportional biasing and linear ranking.

well reflected in the proportional biasing based solely on fitness because 0.96 is not much more likely to be selected that the 0.95 even though it represents a significant improvement. To alleviate this problem, the sets were first sorted in order of fitness. Then, a linear ranking probability, [25]

$$P(i) = \frac{2(N+1-i)}{N(N+1)} \tag{1}$$

was used to assign every strings probability of being selected for a genetic operator. Here, N is the number of sets of features (20 in this case) and i is the string in question $(1 \ldots N)$. This allows much more probability separation between those sets that are very close in A_z because rank, not fitness, is used to determine the probability of being selected. The results of the genetic algorithm runs using this selection rule are shown in Figure 3 (labeled "Linear Ranking"). Notice that the overall performance is dramatically improved. The random search results are displayed to serve as a baseline for evaluating performance.

5.2.2 Results to Date

Segmenting lesions is a vital step in many computerized mass-detection schemes for digital (or digitized) mammograms. In order to improve the classification ability of extracted features, we enhanced the lesion segmentation algorithm. We have

developed two novel lesion segmentation techniques—one based on a single feature called the radial gradient index (*RGI*) and one based on simple probabilistic models to segment mass lesions, or other similar nodular structures, from surrounding background [30]. In both methods a series of image partitions is created using gray-level information as well as prior knowledge of the shape of typical mass lesions. With the former method the partition that maximizes the *RGI* is selected. In the latter method, probability distributions for gray-levels inside and outside the partitions are estimated, and subsequently used to determine the probability that the image occurred for each given partition. The partition that maximizes this probability is selected as the final lesion partition (contour). We tested these methods against a conventional region-growing algorithm using a database of biopsy-proven, malignant lesions and found that the new lesion segmentation algorithms more closely match radiologists' outlines of these lesions. At an overlap threshold of 0.30, gray level region growing correctly delineates 62% of the lesions in our database while the *RGI* and probabilistic segmentation algorithms correctly segment 92% and 96% of the lesions, respectively.

It is well understood that binary classifiers have two implicit objective functions describing their performance. Traditional methods of classifier training attempt to combine these two objective functions into one, so that conventional scalar optimization techniques can be utilized. This involves incorporating *a priori* information into the aggregation method so that the resulting performance of the classifier is satisfactory for the task at hand. We have investigated the use of a niched Pareto multiobjective genetic algorithm for classifier optimization [31]. With niched Pareto genetic algorithms, an objective vector is optimized instead of a scalar function, eliminating the need to aggregate classification objective functions. The niched Pareto genetic algorithm returns a set of optimal solutions that are equivalent in the absence of any information regarding the preferences of the objectives. The *a priori* knowledge that was used for aggregating the objective functions in conventional classifier training can instead be applied post-optimization to select from one of the series of solutions returned from the multiobjective genetic optimization. We have applied this technique to train a linear classifier and an artificial neural network using simulated datasets. The performances of the solutions returned from the multiobjective genetic optimization represent a series of optimal (sensitivity, specificity) pairs, which can be thought of as operating points on an ROC curve. All possible ROC curves for a given dataset and classifier are less than or equal to the ROC curve generated by the niched Pareto genetic optimization.

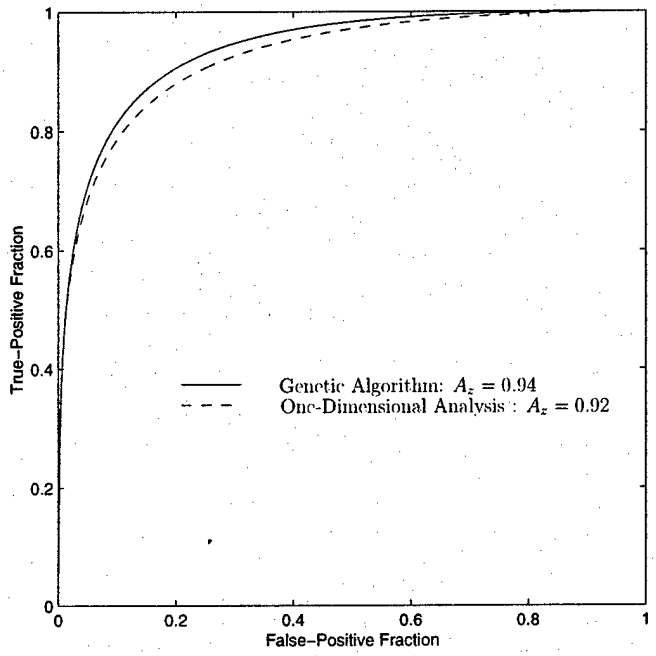

Figure 4: Maximum round-robin ROC curves for the genetically selected features and the features selected using one-dimensional analysis.

5.2.3 Comparison of Genetic Algorithm with Other Techniques

In order to accurately determine how well the genetic algorithm is performing it must be rigorously tested against other methods of selecting or searching for subset of features. Prior research [17] incorporated a one- dimension separation method for determining the subset of features to be used in the ANN. A preliminary comparison between one-dimensional analysis and the performance of the genetic algorithm is shown in Figures 4 and 5. The maximum round-robin A_z increased from 0.92 to 0.94. As shown in Figure 5, at a sensitivity of 89%, there are about 4 fewer false-positive per image using the genetic algorithm over the previous feature selection method. This represents a substantial improvement. This preliminary comparison exhibits the improvements which are possible with genetic algorithms but to fully test the GA it must be compared with other search techniques such as simulated annealing, algebraic techniques using discrete derivatives and more.

5.2.4 Results to Date

We have investigated various methods of feature selection for two different data classifiers used in the computerized detection of mass lesions in digital mammograms

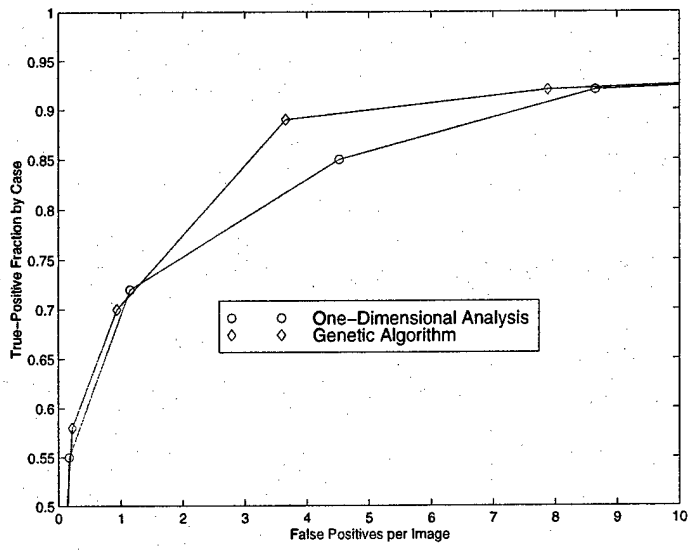

Figure 5: Comparison between the FROC from the genetic algorithm and the FROC using one-dimensional feature separation analysis.

[32]. Numerous features were extracted from abnormal and normal breast regions from a database consisting of 210 individual mammograms. A stepwise method, a genetic algorithm and individual feature analysis were employed to select a subset of features to be used with linear discriminants. Similar techniques were also employed for an artificial neural network classifier. In both tests the genetic algorithm was able to either outperform or equal the performance of other methods.

Table 1 shows the A_z values for the feature selection methods used for determining the inputs for a linear discriminant. Wilks' lambdas are also shown. It is clear from the table that selecting features based on their individual performance is inadequate. In Figure 6 the three different feature selection methods are compared using the ROC curves when 9 features are selected by each method. The A_z values for the feature sets selected by the genetic algorithm and the stepwise method are statistically significantly $p < 0.05$) better than that of the single feature analysis method. The genetic algorithm shows a slight advantage over the stepwise selection method but it is not statistically significant ($p = 0.23$).

Table 2 shows preliminary results from the ANN feature selection methods. It should be noted that multiple genetic algorithm runs were required meaning that the genetic algorithm did have trouble with local maxima. This might suggest that the probability of mutation be increased, as well as the population size, to allow for more diversity throughout the runs. As the table shows the set of features selected by the genetic algorithm was able to outperform the other two methods but the results were

Method	A_z	Wilks' Lambda	Number of Features
	0.93	0.53	9
Single Feature	0.92	0.53	10
Analysis	0.93	0.51	11
	0.94	0.50	12
Stepwise	0.94	0.47	9
	0.95	0.47	9
Genetic	0.95	0.47	10
Algorithm	0.95	0.46	11
	0.95	0.46	12

Table 1: Summary of results from the feature selection methods for linear discriminants.

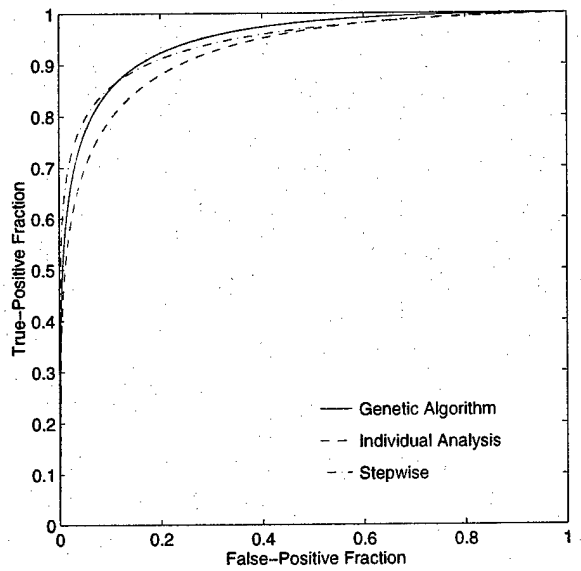

Figure 6: ROC curves for the three different linear discriminant features selection methods when 9 features were selected by each.

Method	Cross Validation A_z	Number of Features
Single Feature Analysis	0.96	11
Forward Selection	0.97	11
Genetic Algorithm	0.98	10

Table 2: Summary of results from the feature selection methods for artificial neural networks.

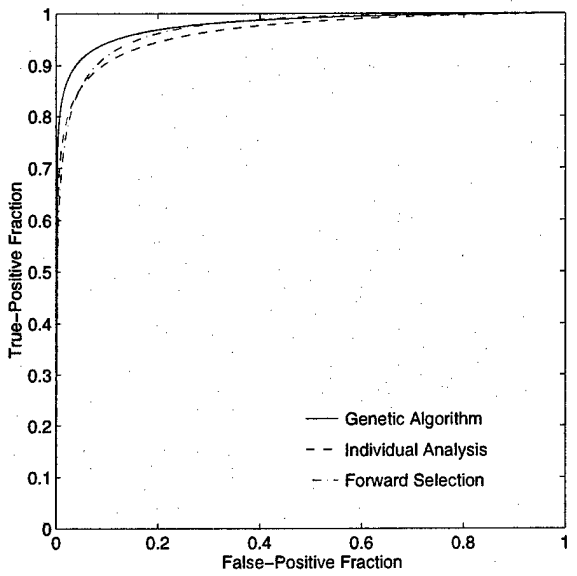

Figure 7: Cross validation ROC curves for ANN feature selectors.

not statistically significant ($p = 0.06$ for the individual analysis selector and $p = 0.15$ for the forward selector). The corresponding ROC curves are shown in Figure 7.

5.2.5 Analysis of Selected Features

One of the biggest mistakes that can be made with powerful search techniques is that the results could be taken and used without study. In order to use the GA properly, it is necessary to do some extensive studies on the features that the GA selects for inputs. Many CAD schemes have selected their features based on what radiologist's say they look for when analyzing images. The GA approach is different; it takes many features and selects the few that combine to perform the best. It is vital that the two sets of features are compared. This will provide two things: First, it could serve as validation for the GA's selected features. If the GA consistently

	Feature	Neighborhood	Orientation
*	contrast deviation	–	–
*	average vertical gradient	margin	–
	average lower 50% along the radial direction	margin	radial
	minimum	grown	Cartesian
	minimum	grown	radial
	full width at half maximum	grown	radial
	standard deviation	grown	Cartesian
	height	periphery	Cartesian
	average lower 50% along the radial direction	periphery	Cartesian
*	standard deviation	ROI	radial

Table 3: Final features selected from 10 genetic algorithm runs. Starred features were selected using the previous one-dimensional analysis as well.

selects features that radiologist's use then we are confident that it is performing as it should. Second, it may help others, namely radiologists, gain insight into other aspects of the image that they might want to look at if the GA selects additional features previously unexpected.

One possible method for obtaining this information involves having experienced radiologists study the images in the ANN training database and rate the visual features that they believe they used to ascertain whether or not an area was a mass or not. If they did indeed correctly classify the area as a mass then it may be possible to gain insight into their classification techniques as well as compare those features deemed important by them with those features selected by the genetic algorithm.

5.2.6 Results to Date

In order to confirm that the genetic algorithm is performing as expected, the features selected by the GA must be analyzed to ensure that they have physical meaning [32,33]. This is a difficult task because the GA selects features that perform well in combination and thus the utility of analyzing features one at a time may be obscured. Table 3 lists the final set of features selected by the GA analysis. The first feature selected was the contrast deviation. This feature was also selected by the previous one-dimensional feature selection method. Actual lesions tend to have a gradual pixel gradient as one moves further away from the center of the lesion. Because of the difficulties in region-growing, false-positives tend to have more uniform centers. Thus, lesions will actually have a larger variation of contrasts than false-positives. The rest of the features selected by the GA are gradient-based. Two features were

selected from the margin neighborhood. The average magnitude of the vertical gradient along the margin measures the sharpness of the borders. Lesions tend to have sharper margins than false-positives so they will generally have larger average gradients along the margin. The second margin feature measured very similar properties. The average of the lower 50% of the magnitude of the gradients projected onto the radial axis measures sharpness and circularity. Very distinct circular lesions will have a large value for this feature while lesions that are either non-circular or have indistinct borders will have smaller values of this feature. In the grown neighborhood, features which measure similar properties were selected. The minimum value of both the x-axis gradient-weighted and the radial gradient-weighted histograms were selected. A lesion, which is more likely to be circular than a false-positive, will have a high minimum gradient value on the x-axis gradient-weighted histogram. Conversely, it will have a very low minimum on the radial gradient-weighted histogram. One would also expect a circular lesion to have a relatively flat gradient-weighted histogram so the standard deviation of the histogram values would be small. This is the reason the standard deviation of the grown region was selected by the GA. Again, features that measure similar properties in a different manner have been selected. The GA also selected the full-width half-maximum of the radial gradient-weighted histogram which is a measure of spiculation [34]. A feature having to do with both spiculation and edge sharpness is the height of the gradient-weighted histogram in the periphery neighborhood. In the periphery the average lower 50% of the radial gradients was selected for the same reason this feature was selected in the margin neighborhood. Finally, within the entire ROI, the standard deviation of the radial gradient histogram was selected because of the larger gradients present in true lesions. This features is similar to the measurement of the height of the histograms which indicates both circularity and edge sharpness.

5.2.7 Development of Parallel Genetic Algorithm

The Kurt Rossmann Laboratories recently acquired the services of Argonne National Laboratory's Supercomputing Center. This provides the department access to a 128-node IBM Scalable Power Parallel SP1/SP2 system. The 128 processors allow up to 128 programs to run in parallel while sharing information without any effect on the speed of each program. Because of the complexity of running a genetic algorithm with a high cardinality and with variable length genes, such as the one needed for the optimization of ANN inputs, performance may be greatly improved by running many genetic algorithms in parallel and sharing information at specific times.

Previous research in parallel genetic algorithms has shown that multiple genetic algorithms running in parallel can provide dramatic improvements in GA performance [35].

5.2.8 Results to Date

Development continues to incorporate the genetic algorithms developed at the Kurt Rossmann Laboratories into the parallel GA packages developed at Argonne National Laboratory.

6 Conclusions

We have developed two new methods of seeded lesion segmentation for use in digital mammography. These new methods substantially outperform conventional region growing segmentation. At an overlap threshold of 0.3, region growing correctly identified 62% of the lesions in our database, while the *RGI*-based and probabilistic segmentation methods correctly segmented 92% and 96% of the lesions, respectively. With these new segmentation results we hope to find and extract new features that will help differentiate between actual lesions and false detections, thus improving the overall performance of computerized mass detection.

We have studied the use of a niched Pareto genetic algorithm in training two popular diagnostic classifiers. Unlike conventional classifier training techniques that formulate the problem as the solution to a scalar optimization, the NP-GA explicitly addresses the multiobjective nature of the training task. It has been demonstrated that the multiobjective approach removes the ambiguity associated with defining a scalar measure of classifier performance, and that it returns a set of optimal solutions that are equivalent in the absence of any information regarding the preference of the objectives (sensitivity, specificity). The performances of these solutions can be interpreted as operating points on an optimal ROC curve, describing the limiting tradeoffs between sensitivity and specificity that are achievable by that classifier, given the available training data. The task of classifier optimization and ROC curve generation are combined into a single task. It was demonstrated that constructing the ROC curve in this way may result in a better ROC curve than is produced by conventional methods of ROC curve generation. The NP-GA optimization typically requires more computation time than do conventional non-stochastic optimization methods, which may limit its application to certain problems. The advantages of the NP-GA approach to classifier training become amplified when the number of classes to be classified increases beyond two.

We have introduced feature selection methods and compare their utility with two different classifiers. The results from the linear discriminant analysis show that the genetic algorithm feature selection method is as good if not better than the stepwise method. Similar results were obtained for the artificial neural network classifiers but the results were not as strong. As with all studies employing neural networks, it is possible that there is over-fitting of the data. We attempted to minimize this effect by simplifying the structure of our networks and by employing cross validation or leave-one-out tests. Future work will include investigations performed on larger data sets.

We will continue to develop and test the enhanced segmentation algorithms, the features extracted using these new segmentation algorithms, feature selection methods

(namely GA feature selection) and MOGA classifier optimization for use in digital mammography. We will also continue to pursue the development of parallel GAs with Argonne National Laboratories.

7 Papers and Presentations

The following papers have been submitted to peer review journals:

- "Automated Seeded Lesion Segmentation on Digital Mammograms." Matthew A. Kupinski and Maryellen L. Giger. Submitted and accepted by *IEEE Transactions on Medical Imaging.* [30]
- "Multiobjective Genetic Optimization of Diagnostic Classifiers with Implications for Generating ROC Curves." Matthew A. Kupinski and Mark A. Anastasio. Submitted and under review by *IEEE Transactions on Medical Imaging.* [31]
- "Optimization and FROC Analysis of Rule-Based Detection Schemes Using a Multiobjective Approach." Mark A. Anastasio, Matthew A. Kupinski and Robert M. Nishikawa. Submitted as a short paper and under review by *IEEE Transactions on Medical Imaging.* [36]

The following proceeding papers have been published:

- "Optimization of Neural Network Inputs with Genetic Algorithms." Matthew A. Kupinski, Maryellen L. Giger and Kunio Doi. "Digital Mammography 96: Proceedings of the 3rd International Workshop on Digital Mammography." Chicago, June 1996. [33]
- "Feature Selection and Classifiers for the Computerized Detection of Mass Lesions in Digital Mammography." Matthew A. Kupinski and Maryellen L. Giger. *IEEE International Congress on Neural Networks* June 1997. [32]
- "Investigation of Regularized Neural Networks for the Computerized Detection of Mass Lesions in Digital Mammograms." Matthew A. Kupinski and Maryellen L. Giger. *IEEE Engineering in Medicine and Biology Society Conference* October 1997. [37]
- "Optimization of Computer-Aided Diagnosis Schemes Using a Multiobjective Approach." Mark A. Anastasio, Matthew A. Kupinski, Robert M. Nishikawa and Maryellen L. Giger. *IEEE Medical Imaging Conference* 1998 (in press). [38]

The following presentations have been given:

- "Use of Genetic Algorithms in the Computerized Detection of Masses in Digital Mammograms." Matthew A. Kupinski and Maryellen L. Giger. *American Association of Physicists in Medicine* 1996.

- "Optimization of Neural Network Inputs with Genetic Algorithms." Matthew A. Kupinski, Maryellen L. Giger and Kunio Doi. "Digital Mammography 96: Proceedings of the 3rd International Workshop on Digital Mammography." Chicago, June 1996.
- "Neural Network Model Selection in the Computerized Detection of Mass Lesions in Digital Mammograms." Matthew A. Kupinski and Maryellen L. Giger. *American Association of Physicists in Medicine* 1997.
- "Feature Selection and Classifiers for the Computerized Detection of Mass Lesions in Digital Mammography." Matthew A. Kupinski and Maryellen L. Giger. *IEEE International Congress on Neural Networks* June 1997.
- "Investigation of Regularized Neural Networks for the Computerized Detection of Mass Lesions in Digital Mammograms." Matthew A. Kupinski and Maryellen L. Giger. *IEEE Engineering in Medicine and Biology Society Conference* October 1997.
- "Probabilistic Lesion Segmentation in Digital Mammography." Matthew A. Kupinski and Maryellen L. Giger. *American Association of Physicists in Medicine* 1998.
- "Optimization of Computer-Aided Diagnosis Schemes Using a Multiobjective Approach." Mark A. Anastasio, Matthew A. Kupinski, Robert M. Nishikawa and Maryellen L. Giger. *IEEE Medical Imaging Conference* 1998.

References

[1] E. Silverberg, C. C. Boring, and T. S. Squires, *Cancer Statistics*, vol. 40. 1990.

[2] L. W. Basset and R. H. Gold, *Breast Cancer Detection. Mammography and Other Methods in Breast Imaging.* Grune and Stratton, 1987.

[3] J. Lissner, M. Kessler, and G. Anhalt, "Developments in methods for early detection of breast cancer," in *Early Breast Cancer* (J. Aandler and J. Baltzer, eds.), (Berlin), Springer-Verlag, 1984.

[4] I. Andersson, "What can we learn from interval carcinomas?," *Recent Results in Cancer Research*, vol. 90, pp. 191–193, 1984.

[5] C. J. Baines, A. B. Miller, and C. Wall, "Sensitivity and specificity of first screen mammography in the canadian national breast screening study. a preliminary report from five centers," *Radiology*, vol. 160, pp. 295–298, 1986.

[6] J. E. Martin, M. Moskowitz, and J. R. Milbrath, "Breast cancers missed by mammography," *American Journal of Radiology*, vol. 132, pp. 737–739, 1979.

[7] S. R. Pollei, F. A. Mettler, S. A. Bartow, G. Moradian, and M. Moskowitz, "Occult breast cancer. prevalence and radiographic detectability," *Radiology*, vol. 16, pp. 459–462, 1987.

[8] J. B. Buchanan, J. S. Spratt, and L. S. Heuser, "Tumor growth, doubling times, and the inability of the radiologist to diagnose certain cancers," *Radiologic Clinics of North America*, vol. 21, pp. 115–126, 1983.

[9] T. Holland, M. Mrvunac, J. H. C. L. Hendrirks, and B. Bekker, "So-called interval cancers of the breast. pathologic and radiographic analysis," *Cancer*, vol. 49, pp. 2527–2533, 1982.

[10] E. L. Thurfjell, K. A. Lernevall, and A. A. Taube, "Benefit of independent double reading in a population-based mammography screening program," *Radiology*, vol. 191, pp. 241–244, 1994.

[11] C. E. Metz and J.-H. Shen, "Gains in accuracy from replicated readings of diagnostic images: Prediction and assessment in terms of roc analysis," *Medical Decision Making*, vol. 12, pp. 60–75, 1992.

[12] F.-F. Yin, M. L. Giger, C. J. Vyborny, K. Doi, and R. A. Schmidt, "Comparison of bilateral-subtraction and single-image processing techniques in the computerized detection of mammographic masses," *Investigative Radiology*, vol. 28, pp. 473–481, 1993.

[13] F.-F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique," *Medical Physics*, vol. 21, pp. 445–452, 1994.

[14] F.-F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," *Medical Physics*, vol. 18, pp. 955–963, 1991.

[15] K. Doi, M. L. Giger, R. M. Nishikawa, K. R. Hoffmann, H. MacMahon, R. A. Schmidt, and K.-G. Chua, "Digital radiography: A useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images," *Acta Radiologica*, vol. 34, pp. 426–439, 1993.

[16] M. L. Giger, R. M. Nishikawa, K. Doi, F.-F. Yin, C. J. Vyborny, R. A. Schmidt, C. E. Metz, Y. Wu, H. MacMahon, and H. Yoshimura, "Development of a "smart" workstation for use in mammography," in *SPIE*, vol. 1445, pp. 101–103, 1991.

[17] M. Kupinski, M. L. Giger, P. Lu, and Z. Huo, "Computerized detection of mammographic lesions: Performance of artificial neural network with enhanced feature extraction," in *SPIE*, vol. 2434, pp. 598–605, 1995.

[18] Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancers," *Radiology*, vol. 187, pp. 81–87, 1993.

[19] Y. Wu, M. L. Giger, K. Doi, C. E. Metz, R. M. Nishikawa, C. J. Vyborny, and R. A. Schmidt, "Application of artificial neural networks in mammography for the diagnosis of breast cancer," in *SPIE*, vol. 1778, pp. 19–27, 1992.

[20] F.-F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Investigation of feature-analysis techniques," *Journal of Digital Imaging*, vol. 7, pp. 18–26, 1994.

[21] Y.-H. Pao, *Adaptive Pattern Recognition and Neural Networks*. Addison-Wesley, 1989.

[22] K. A. D. Jong, "Introduction to the second special issue on genetic algorithms," *Machine Learning*, vol. 5, pp. 351–353, 1990.

[23] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company, Inc., 1989.

[24] P. Sutton and S. Boyden, "Genetic algorithms: A general search procedure," *American Journal of Physics*, vol. 62, pp. 549–552, 1994.

[25] A. J. Katz and P. R. Thrift, "Generating image filters for targe recognition by genetic learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 906–910, 1994.

[26] G. Roth and M. D. Levine, "Geometric primitive extraction using a genetic algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 901–905, 1994.

[27] C. E. Metz, "Basic principles of roc analysis," *Seminars in Nuclear Medicine*, vol. VIII, pp. 283–298, 1978.

[28] C. E. Metz, "Roc methodology in radiologic imaging," *Investigative Radiology*, vol. 21, pp. 720–733, 1986.

[29] C. E. Metz, J.-H. Shen, and B. A. Herman, "New methods for estimating a binormal roc curve from continuously-distributed test results," in *Joint Meeting of the American Statistical Society and the Biometric Society*, 1990.

[30] M. A. Kupinski and M. A. Anastasio, "Automated seeded lesion segmentation on digital mammograms," *IEEE Transactions on Medical Imaging*, 1998 (in press).

[31] M. A. Kupinski and M. A. Anastasio, "Multiobjective genetic optimization of diagnostic classifiers with implications for generating ROC curves," *IEEE Transactions on Medical Imaging*, 1998 (in review).

[32] M. A. Kupinski and M. L. Giger, "Feature selection and classifiers for the computerized detection of mass lesions in digital mammography," in *Proceedings of the 1997 International Conference on Neural Networks (ICNN '97)*, (Houston, TX), pp. 1336–1339, IEEE/ICNN, June 9-12 1997.

[33] M. A. Kupinski, M. L. Giger, and K. Doi, "Optimization of neural network inputs with genetic algorithms," in *Digital Mammography* (K. Doi, ed.), International Congress Series, pp. 401–404, Elsevier, 1996.

[34] Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, and P. Lu, "Analysis of spiculation in the computerized classification of mammographic masses," *Medical Physics*, vol. 22, pp. 1569–1579, 1995.

[35] D. Levine, "A parallel genetic algorithm for the set partitioning problem," *ANL-94-23*, 1994.

[36] M. A. Kupinski and M. A. Anastasio, "Optimization and froc analysis of rule-based detection schemes using a multiobjective approach," *IEEE Transactions on Medical Imaging*, 1998 (in review).

[37] M. A. Kupinski and M. L. Giger, "Investigation of regularized neural networks for the computerized detection of mass lesions in digital mammograms," in *Proceedings of the 19th International Conference of Engineering in Medicine and Biology*, (Chicago, IL), pp. 1336–1339, IEEE/EMBS, Oct. 30-Nov. 2 1997.

[38] M. A. Anastasio, M. A. Kupinski, R. M. Nishikawa, and M. L. Giger, "Optimization of computer-aided diagnosis schemes using a multiobjective approach," in *Proceedings of the 1998 IEEE Medical Imaging Conference*, IEEE, 1998 (in press).